Palmer, S., Manns, S., Cramp, F., Lewis, R., & Clark, E. M. (2017). Test-retest reliability and smallest detectable change of the Bristol Impact of Hypermobility (BIoH) questionnaire. *Musculoskeletal Science and Practice*, *32*, 64-69. https://doi.org/10.1016/j.msksp.2017.08.007

TITLE PAGE

Title: Test-retest reliability and smallest detectable change of the Bristol Impact of Hypermobility (BIoH) questionnaire

Author Names & Affiliations: Palmer S[a], Manns S[a], Cramp F[a], Lewis R[b], Clark EM[c]

[a]Department of Allied Health Professions, University of the West of England, Bristol, BS16 1DD, UK.

[b]Department of Physiotherapy, North Bristol NHS Trust, Southmead Hospital, Westbury-on-Trym, Bristol, BS10 5NB, UK.

[c]Musculoskeletal Research Unit, University of Bristol, Southmead Hospital, Westbury-on-Trym, Bristol, BS10 5NB, UK.

Corresponding author: Professor Shea Palmer, Department of Allied Health Professions, University of the West of England, Blackberry Hill, Bristol, UK, BS10 1DD. Tel: +44 (0)117 3288919. Email: Shea.Palmer@uwe.ac.uk. ORCID ID: orcid.org/0000-0002-5190-3264,

Keywords: Benign hypermobility syndrome; Ehlers-Danlos Syndrome, Hypermobility Type; Test-retest reliability; Questionnaire

Word count: 3,209

ABSTRACT

Objective: The Bristol Impact of Hypermobility (BIoH) questionnaire is a patient-reported outcome measure developed in conjunction with adults with Joint Hypermobility Syndrome (JHS). It has demonstrated strong concurrent validity with the Short Form-36 (SF-36) physical component score but other psychometric properties have yet to be established. This study aimed to determine its test-retest reliability and smallest detectable change (SDC).

Design: A test-retest reliability study.

Setting: Participants were recruited from the Hypermobility Syndromes Association, a patient organisation in the United Kingdom.

Patients: Recruitment packs were sent to 1,080 adults who had given permission to be contacted about research.

Main Outcome Measures: BIoH and SF-36 questionnaires were administered at baseline and repeated two weeks later. An 11-point global rating of change scale (-5 to +5) was also administered at two weeks. Test-retest analysis and calculation of the SDC was conducted on 'stable' patients (defined as global rating of change -1 to +1).

Results: 462 responses were received. 233 patients reported a 'stable' condition and were included in analysis (95% women; mean (SD) age 44.5 (13.9) years; BIoH score 223.6 (54.0)). The BIoH questionnaire demonstrated excellent test-retest reliability (ICC 0.923, 95% CI 0.900-0.940). The SDC was 42 points (equivalent to 19% of the mean baseline score). The SF-36 physical and mental component scores demonstrated poorer test-retest reliability and larger SDCs (as a proportion of the mean baseline scores).

Conclusion: The results provide further evidence of the potential of the BIoH questionnaire to underpin research and clinical practice for people with JHS.

MANUSCRIPT

INTRODUCTION

Joint Hypermobility Syndrome (JHS) is a heritable connective tissue disorder characterised by excessive joint range of motion and pain (Grahame, 2003). It was previously widely recognised that there was a lack of distinction in the clinical presentation of JHS and Ehlers-Danlos Syndrome, Hypermobility Type (EDS-HT) (Tinkle et al., 2009) and many authors considered them to be the same condition. Indeed recent revision of the classification of EDS has created a more specific diagnostic category to replace both of those terms, this being 'hypermobile EDS' (hEDS) (Malfait et al., 2017). Where patients have symptomatic joint hypermobility but do not meet the diagnostic criteria for other syndromes, the term 'Hypermobility Spectrum Disorder' (HSD) has been adopted (Castori et al., 2017). As the research described in this paper pre-dated the new criteria the term JHS will be used.

There is a distinct lack of good epidemiological data on the prevalence of JHS, although it is likely to be high in musculoskeletal services. For example 30% of a convenience sample of 150 patients screened in a musculoskeletal triage service in the United Kingdom (UK) (Connelly et al., 2015) met the Brighton diagnostic criteria (Grahame et al., 2000). It should be acknowledged, however, that JHS may not have been the reason for referral and it is unlikely that many of these patients would have ordinarily received a diagnosis of JHS. The condition is associated with a wide range of problems including pain, fatigue, reduced proprioception and repeated cycles of injury (Terry et al., 2015). The psychological impact of JHS is also well recognised, including agoraphobia, anxiety, catastrophisation, depression, fear and panic disorders (Terry et al., 2015; Smith et al., 2014a).

Physiotherapy, particularly exercise, is a mainstay of treatment for JHS (Hakim and Grahame, 2004; Keer and Grahame, 2003; Simmonds and Keer, 2007; Simmonds and Keer, 2008; Tinkle, 2008; Keer and Simmonds, 2011). The research evidence for the effectiveness of therapy is inconclusive however, with a limited number of low quality studies reported in the literature (Palmer et al., 2014; Smith et al., 2014b). A recent pilot randomised controlled trial (RCT) of a complex physiotherapy intervention demonstrated positive effects on a range of clinical outcomes (Palmer et al., 2016), although such effects have yet to be confirmed in an adequately powered clinical trial.

A survey of physiotherapy practice in the UK (Palmer et al., 2015) suggested that there was a mismatch between what physiotherapists considered to be the aims of physiotherapy for JHS and the outcome measures that they used to assess the effectiveness of management. A condition-specific outcome measure has therefore recently been developed in close collaboration with adults with JHS in an attempt to more adequately capture the wide-ranging impact of the condition (Palmer et al., 2017). The 'Bristol Impact of Hypermobility' (BIoH) questionnaire gives a maximum score of 360, with higher scores representing more severe impact of the condition on the person's life. It has strong concurrent validity (r=-0.725, n=615) with the physical component score of an established general health questionnaire, the Short Form-36 (SF-36) (Palmer et al., 2017). The SF-36 was chosen as a comparator as it has previously been demonstrated to be sensitive to change following an exercise intervention (Ferrell et al., 2004). Although the BIoH questionnaire has demonstrated initial promise, it needs further evaluation before it can be used confidently to support future research and clinical practice in this area. Adequate test-retest reliability is an important psychometric property for an outcome measure, ensuring that scores are constant over time in those patients who report that their condition is stable (Polit, 2014).

This project therefore aimed to conduct a test-retest reliability evaluation of the BIoH questionnaire over a two-week period and to calculate the smallest detectable change (SDC). Secondary aims were to conduct a limits of agreement analysis and an exploratory sensitivity to change analysis. Findings were compared with the SF-36.

METHODS

This study received a favourable opinion from the Faculty of Health & Applied Sciences Ethics Sub-Committee, University of the West of England, Bristol, UK (HAS/15/01/99).

Recruitment

Recruitment packs were distributed to 1,080 adult members of the Hypermobility Syndromes Association (HMSA) who had given permission to be contacted about research. The HMSA is a

patient organisation in the UK. Packs included a letter of invitation, information sheet, consent form, a brief demographic and screening questionnaire, and copies of the BIoH and SF-36v2 (OptumInsight Life Sciences Inc.) questionnaires. The information sheet included an invitation to contact the research team for additional information if required. Those willing to take part were asked to complete and return the consent form and questionnaires using a pre-paid envelope. Completion of the BIoH questionnaire takes approximately 10 minutes (Palmer et al., 2017) and the SF-36 7-10 minutes (Coons et al., 2000). The inclusion criteria were: ≥16 years old; fulfilled two or more JHS screening questions (Hakim and Grahame, 2003) and/or received a formal diagnosis (by a healthcare professional) of JHS or EDS-HT; no other formally diagnosed conditions affecting physical function; written informed consent; complete baseline and follow-up BIoH questionnaire data. The JHS screening questionnaire has previously demonstrated 84% sensitivity and 80-89% specificity (Hakim and Grahame, 2003).

Two weeks later repeat BIoH and SF-36 questionnaires were sent to consenting participants to complete independently, along with a global rating of change (GRC) scale (Kamper et al., 2009). A retest period of one to two weeks is typical for many test-retest studies (Polit, 2014). To allow sufficient time for postal administration of the questionnaires and because the BIoH asks about the previous seven days, a period of two weeks was chosen. The GRC scale was an 11-point numerical rating scale which asked patients *"With respect to your joint hypermobility, how would you describe yourself now compared with two weeks ago when you last completed this questionnaire (please circle a number)?"* Anchors were at -5 (*"Very much worse"*), 0 (*"Unchanged"*) and +5 (*"Very much better"*).

No formal sample size calculation was conducted, although a minimum of 50 participants is recommended for limits of agreement analysis (Altman, 1991; Terwee et al., 2007). It was anticipated that many more participants would be recruited as a 42% response rate from HMSA members was previously achieved when initially developing the BIoH questionnaire (Palmer et al., 2017).

Data analysis

Data analysis was conducted using IBM SPSS Statistics (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp). The BIoH questionnaire scoring

guidelines were followed to calculate BIoH scores (Palmer et al., 2017). SF-36 scoring software (Version 4.5. Lincoln, RI: OptumInsight Life Sciences Inc.) was used to calculate SF-36v2 component scores. Questionnaire data was manually entered into an SPSS spreadsheet (SM, PhD) and the accuracy of data entry was audited and verified by a second researcher (SP, PhD).

The same analysis procedures were used for the BIoH and SF-36 questionnaires. Test-retest reliability analysis was conducted on 'stable' patients (defined as those who scored -1 to +1 on the GRC score). This was to ensure that assessment of test-retest reliability was not contaminated by fluctuations in patients' conditions, a common feature of JHS (Terry et al., 2015). A change of plus or minus one point on an 11-point GRC scale is considered to be clinically unimportant (Kamper et al., 2009). Descriptive statistics were used to compare the characteristics of 'stable' patients against the wider sample of patients who met the inclusion criteria. Intraclass Correlation Coefficient (ICC) (two-way mixed effects, single measure, absolute agreement) (Koo and Li, 2016), standard error of measurement (SEM) for agreement and SDC were calculated for the BIoH questionnaire and SF-36 scores (Polit, 2014). Details for calculation of the SEM and SDC are available in Table 3. The 95% limits of agreement were plotted for the BIoH scores (Bland and Altman, 1986).

An exploratory sensitivity to change analysis was conducted in all participants who met the inclusion criteria. A scatterplot explored the relationship between GRC score and change in BIoH score. Pearson's correlation coefficients further explored the relationships between GRC scores and change in BIoH and SF-36 scores.

RESULTS

462 responses were received from HMSA members (43% response rate). 99 were excluded on the following basis: 82 failed to return week two questionnaires; Seven had incomplete follow-up data; Five had been diagnosed with another connective tissue disease; Three were <16 years; One could not be diagnosed with JHS on the basis of the screening questions; One had incomplete baseline data. A total of 363 participants therefore met the inclusion criteria and completed baseline and follow-up BIoH questionnaires. Of these, 233 scored between -1 and +1 on the GRC score across the two-

week study period and thus were classified as ‘stable’ for the purposes of the test-retest reliability analysis. Kolmogorov-Smirnov tests confirmed that data for the BIoH questionnaire and SF-36 PCS conformed to a normal distribution at both baseline and two-weeks (n=233). Although the SF-36 MCS data deviated from normality (n=233), a pragmatic decision was made to use parametric analysis throughout as the choice of statistical approach for analysis of SF-36 data in people with chronic pain has previously been found to have no effect on the outcome (Torrance et al 2009).

Table 1 illustrates that there were no systematic differences between ‘stable’ participants (n=233) and the larger sample of participants who met the inclusion criteria (n=363), suggesting that the ‘stable’ patients were representative of the larger group from which they came. Table 2 presents information relating to the diagnostic make-up of the wider sample (n=363).

The distribution of GRC scores across the full data set (n=363) can be seen in Figure 1. ‘No change’ was defined as a GRC score from -1 to +1 (n=233). An ICC value of 0.923 (95% Confidence Interval (CI) 0.900-0.940) was calculated for the BIoH questionnaire, indicating excellent test-retest reliability (n=233). Corresponding values for the SF-36 Physical Component Score (PCS) and Mental Component Score (MCS) were ICC 0.887 (95% CI 0.855-0.912) and ICC 0.778 (95% CI 0.721-0.825) respectively (n=233).

Calculations relevant to limits of agreement analysis are contained in Table 3. The mean difference in BIoH scores across the two weeks for ‘stable’ patients (n=233) was -3.26 points, suggesting a trend towards slightly lower scores on repeat testing. The 95% CI for $\bar{d}$ for the BIoH questionnaire did not cross zero which confirms slight bias on repeated measurement (in this case a tendency towards a lower score at week two) (Rankin and Stokes, 1998). The SDC (1.96 SD of the difference scores) for the BIoH questionnaire was calculated to be $\pm$42 points (equivalent to 12% of the maximum score (360 points) or 19% of the mean baseline score (223.6 points) observed in the current sample) (n=233). The corresponding values for the SF-36 PCS and MCS were $\pm$9 and $\pm$15 points respectively (equivalent to 9% and 15% of the maximum score (100 points) or 25% and 37% of the mean baseline scores (33.6 and 41.8 points) respectively) (n=233).

The 95% limits of agreement were calculated as the mean difference ± the SDC. Figure 2 presents the limits of agreement plot for the BIoH questionnaire (Bland and Altman, 1986) (n=233). It is difficult to identify clear trends but there is perhaps slightly greater between-days differences (the difference in score between baseline and week 2) in the region of the mean BIoH scores (the mean baseline score was 223.6). These differences seem to be smaller at either end of the range of between-days mean scores, particularly in the lower range.

The relationship between the GRC score and change in the BIoH questionnaire score across the two-week study period is illustrated in Figure 3 for all participants who met the inclusion criteria (n=363). This demonstrates a moderate negative correlation (Pearson's correlation coefficient r=-0.493, $p<0.001$), indicating that improvement measured by the GRC scale was associated with a decrease in the impact of the condition as measured by the BIoH questionnaire. Corresponding SF-36 values were r=0.186 ($p<0.001$) and r=0.203 ($p<0.001$) for the PCS and MCS respectively, indicating weak associations between GRC and improvement in disability as measured by the SF-36.

DISCUSSION

The BIoH questionnaire has demonstrated excellent test-retest reliability over a two-week period, with an ICC value of 0.923. Test-retest reliability coefficient values in excess of 0.85 have previously been recognised as indicating 'excellent' reliability (Polit, 2014). In the present investigation even the lower boundary of the ICC 95% confidence interval for the BIoH questionnaire was equal to 0.9 so this indicates that the questionnaire performed very strongly in this regard. The PCS of the SF-36 also performed very well with an 'excellent' ICC value of 0.887. The MCS performed less well, with an ICC value of 0.778, although this would still be classified as 'adequate' by many researchers (Polit, 2014). So, both the BIoH questionnaire and SF-36 PCS demonstrated excellent test-retest reliability in the context of this study, with marginally better reliability for the BIoH questionnaire.

The SDC values suggested that a change of 42 points or more on the BIoH questionnaire might be considered beyond measurement error (equivalent to 19% of the mean score observed in the current sample). A change of less than this is therefore unlikely to be important clinically. The corresponding

SDC values for the SF-36 PCS and MCS were equivalent to 25% and 37% of the mean scores respectively. This suggests that the BIoH questionnaire may be more sensitive to change, at least in terms of the relative magnitude of measurement error. However it should be acknowledged that these estimates are only based on the boundaries of error around the change score in 'stable' patients (Wright et al., 2012). The minimum clinically important difference would therefore need to be verified through analysis of sensitivity of change following interventions of known efficacy.

A further preliminary analysis of sensitivity to change was performed by analysing the relationship between GRC scale scores and change in the BIoH questionnaire and SF-36 values across the two-week study period. This suggested that the BIoH performed much better, with correlation values approaching -0.5 compared to values for the PCS and MCS in the region of 0.2. This supports an assumption that the BIoH questionnaire may be more sensitive, although again this needs to be a very tentative interpretation until a more robust evaluation of sensitivity to change is conducted.

A number of different outcome measures have been used in previous JHS clinical trials. For example the patient-reported outcome measures used in previous trials of exercise for adults with JHS included in a recent systematic review (Palmer et al., 2014) were the SF-36 (Ferrell et al., 2004), the Arthritis Impact Measurement Scales 2 (AIMS-2) (Sahin et al., 2008) and a questionnaire developed by Barton and Bird (1996). Of those, only the SF-36 captured improvements following exercise (Ferrell et al., 2004). Only one of the five AIMS-2 subscales improved (Sahin et al., 2008), and no changes were evident in the questionnaire used by Barton and Bird (1996). The SF-36 questionnaire was thus a good candidate against which to compare the BIoH questionnaire (Palmer et al., 2017). A fuller critical evaluation of these studies is available in the systematic review by Palmer et al (2014).

The prevalence of pain was shown to be very high, with participants reporting pain in a mean of 8 out of 10 body areas, a similar prevalence to that previously observed (Palmer et al., 2017). This is perhaps unsurprising as participants in both studies were recruited from the same patient organisation. Nonetheless, it reinforces the observation that the pain experienced by those with JHS is widespread.

Additional information related to the five-point screening questionnaire of Hakim and Grahame (2003) has been gathered, suggesting that some questions might be more discriminative than others (Table 2). For example shoulder or kneecap dislocation was only reported by 38% of respondents. The original developers of the questionnaire found similarly low proportions for positive responses to that question, with only 20% and 38% of their cohorts giving an affirmative answer (Hakim and Grahame, 2003). The questionnaire now forms a supplementary part of the diagnostic criteria for hEDS in cases where the Beighton criteria is one point below the age-related cut-off (Malfait et al., 2017). As such, further evaluation of the psychometric properties of the five-point questionnaire is warranted and could include an evaluation of sensitivity and specificity against the primary diagnostic criteria. Although many experts have historically considered JHS and EDS-HT to be the same condition (Tinkle et al., 2009) it was interesting to note that the diagnostic label of JHS (92%) was more commonly reported by participants in this study than EDS-HT (63%). It is not known how many participants would meet the new diagnostic criteria for hEDS (Malfait et al., 2017).

Strengths and limitations

The sample included in the present investigation was very similar to that included in initial validation of the BIoH questionnaire (Palmer et al., 2017). This is perhaps not surprising as participants were recruited through the same patient organisation. The sample was slightly more ethnically diverse than that previously recruited by Palmer et al (2017), although it could still not be considered representative. For example 87.1% of the UK population described themselves as ‘white’ in the 2011 census (Office for National Statistics, 2013), as opposed to 95.3% of all eligible participants or 93.5% of ‘stable’ participants in the present investigation. The sample also lacked diversity in terms of sex (95% women) and educational attainment was much higher than might be expected (25% reported having a postgraduate degree). Although there is generally a lack of good epidemiological data in this area, Engelbert et al (2004) suggested that the prevalence of symptomatic joint hypermobility in adults might be 3.3% of women and 0.6% of men. This gives a woman: man ratio of approximately 5.5 : 1, as opposed to the much higher 19 : 1 of respondents in the present investigation. This might limit the generalisability of study findings.

Use of the five-point JHS screening questionnaire (Hakim and Grahame, 2003) was a strength of the study design. A limitation of the methodology used in the initial validation of the BIoH questionnaire was that a diagnosis of JHS was self-declared (Palmer et al., 2017). Whilst the present investigation still relied upon self-report in relation to the screening questions, the questionnaire has previously exhibited high sensitivity and specificity for JHS (Hakim and Grahame, 2003). This increases confidence in relation to the diagnostic status of participants, although it cannot be assured for all individuals as sensitivity and specificity are 16% and 11-20% short of perfection respectively.

Conclusion

The present investigation demonstrated that the BIoH questionnaire performed better than the SF-36 PCS and MCS in terms of test-retest reliability. The BIoH questionnaire SDC was also smaller than either SF-36 subscale (as a proportion of the mean baseline scores). In addition, there was a stronger correlation between change in the BIoH score and the GRC score than either the SF-36 PCS or MCS, suggesting that it may be more sensitive to change. However, as already suggested, analysis of sensitivity to change following interventions of known efficacy would be necessary to confirm such speculation. It is clear that, on the basis of the psychometric properties investigated in the present study, the BIoH questionnaire performed very well when compared against the SF-36. The results provide further evidence of the psychometric properties of the BIoH questionnaire, enhancing confidence in its potential use for clinical research and practice.

HIGHLIGHTS

- The BIoH questionnaire demonstrated excellent test-retest reliability.
- The smallest detectable change was calculated to be 42 points.
- The BIoH questionnaire performed better than the SF-36.
- Confidence in using the BIoH questionnaire for research and practice is enhanced.

Acknowledgements: We would like to acknowledge the assistance of the Hypermobility Syndromes Association in recruiting participants and to thank participants for taking part in the research. The

Short-Form 36v2 questionnaire was used under license from OptumInsight Life Sciences Inc (license number QM013824).

Conflict of interest statement: There are no conflicts of interest.

Funding support: The study was supported by a small grant from the Faculty of Health & Applied Sciences, University of the West of England, Bristol, UK.

Ethical Approval: The study received a favourable opinion from the Faculty of Health & Applied Sciences Ethics Sub-Committee at the University of the West of England, Bristol (HAS/15/01/99).

REFERENCES

- Altman DG. Practical statistics for medical research. London: Chapman and Hall; 1991.
- Barton LM, Bird HA. Improving pain by the stabilization of hyperlax joints. J Orthop Rheumatol 1996; 9: 46-51.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; i:307-10.
- Castori M, Tinkle B, Levy H, Grahame R, Malfait F, Hakim A. A framework for the classification of joint hypermobility and related conditions. Am J Med Genet C Semin Med Genet 2017; 175(1): 148-57.
- Connelly E, Hakim A, Davenport S, Simmonds J. A study exploring the prevalence of Joint Hypermobility Syndrome in patients attending a Musculoskeletal Triage Clinic. Physiotherapy Practice and Research 2015; 36(1): 43-53.
- Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. Pharmacoeconomics 2000; 17: 13-35.
- Engelbert RH, Uiterwaal CS, van de Putte E, Helders PJ, Sakkers RJ, van Tintelen P, Bank RA.Pediatric generalized joint hypomobility and musculoskeletal complaints: a new entity? Clinical, biochemical, and osseal characteristics. Pediatrics 2004; 113(4): 714-9.
- Ferrell WR, Tennant N, Sturrock RD, Ashton L, Creed G, Brydson G, Rafferty D. Amelioration of symptoms by enhancement of proprioception in patients with joint hypermobility syndrome. Arthritis Rheum 2004; 50: 3323–8.
- Grahame R, Bird HA, Child A. The revised (Brighton 1998) criteria for the diagnosis of benign joint hypermobility syndrome (BJHS). Journal of Rheumatology 2000; 27: 1777-9.
- Grahame R. Hypermobility and hypermobility syndrome. In: Keer R, Grahame R (Editors) Hypermobility syndrome. Recognition and management for physiotherapists. Edinburgh: Butterworth-Heinemann; 2003. p. 1-14.
- Hakim AJ, Grahame R. A simple questionnaire to detect hypermobility: an adjunct to the assessment of patients with diffuse musculoskeletal pain. International Journal of Clinical Practice 2003; 57: 163-6.
- Hakim AJ, Grahame R. Non-musculoskeletal symptoms in joint hypermobility syndrome. Indirect evidence for autonomic dysfunction? Rheumatology 2004; 43(9): 1194-5.

- Kamper SJ, Maher CG, Mackay G. Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design. J Man Manip Ther 2009; 17(3): 163-70.
- Keer R, Grahame R. Hypermobility syndrome. Recognition and management for physiotherapists. Edinburgh: Butterworth Heinemann; 2003.
- Keer R, Simmonds J. Joint protection and physical rehabilitation of the adult with hypermobility. Current Opinion in Rheumatology 2011; 23: 131-6.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine 2016; 15: 155-63.
- Malfait F, Francomano C, Byers P, Belmont J, Berglund B, Black J, Bloom L, Bowen JM, Brady AF, Burrows NP, Castori M, Cohen H, Colombi M, Demirdas S, De Backer J, De Paepe A, Fournel-Gigleux S, Frank M, Ghali N, Giunta C, Grahame R, Hakim A, Jeunemaitre X, Johnson D, Juul-Kristensen B, Kapferer-Seebacher I, Kazkaz H, Kosho T, Lavallee ME, Levy H, Mendoza-Londono R, Pepin M, Pope FM, Reinstein E, Robert L, Rohrbach M, Sanders L, Sobey GJ, Van Damme T, Vandersteen A, van Mourik C, Voermans N, Wheeldon N, Zschocke J, Tinkle B. The 2017 international classification of the Ehlers-Danlos syndromes. Am J Med Genet Part C Semin Med Genet 2017; 175C: 8-26.
- Office for National Statistics. 2011 Census: Ethnic group, local authorities in the United Kingdom. London: Office for National Statistics; 2013.
- Palmer S, Bailey S, Barker L, Barney L, Elliott A. The effectiveness of therapeutic exercise for joint hypermobility syndrome: a systematic review. Physiotherapy 2014; 100: 220-7.
- Palmer S, Cramp F, Lewis R, Muhammad S, Clark E. Diagnosis, management and assessment of adults with joint hypermobility syndrome: a UK-wide survey of physiotherapy practice. Musculoskeletal Care 2015; 13(2): 101-11.
- Palmer S, Cramp F, Clark E, Lewis R, Brookes S, Hollingworth W, Welton N, Thom H, Terry R, Rimes KS, Horwood J. The feasibility of a randomised controlled trial of physiotherapy for adults with joint hypermobility syndrome. Health Technology Assessment 2016; 20(47).
- Palmer ST, Cramp F, Lewis R, Gould G, Clark EM. Development and initial validation of the Bristol Impact of Hypermobility (BIoH) questionnaire. Physiotherapy 2017, 103(2):186-192.
- Polit DF. Getting serious about test-retest reliability: a critique of retest research and some recommendations. Qual Life Res 2014; 23: 1713-20.

- Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clinical Rehabilitation 1998; 12: 187-99.
- Sahin N, Baskent A, Cakmak A, Salli A, Ugurlu H, Berker E. Evaluation of knee proprioception and effects of proprioception exercise in patients with benign joint hypermobility syndrome. Rheumatol Int 2008; 28: 995-1000.
- Simmonds JV, Keer RJ. Hypermobility and hypermobility syndrome. Manual Therapy 2007; 12(2): 298-309.
- Simmonds JV, Keer RJ. Hypermobility and the hypermobility syndrome. Part 2: Assessment and management of hypermobility syndrome: illustrated via case studies. Manual Therapy 2008; 13(2): e1–11.
- Smith TO, Easton V, Bacon H, Jerman E, Armon K, Poland F, Macgregor AJ. The relationship between benign joint hypermobility syndrome and psychological distress: a systematic review and meta-analysis. Rheumatology (Oxford) 2014a; 53(1): 114-22.
- Smith TO, Bacon H, Jerman E, Easton V, Armon K, Poland F, Macgregor AJ. Physiotherapy and occupational therapy interventions for people with benign joint hypermobility syndrome: a systematic review of clinical trials. Disabil Rehabil 2014b; 36(10): 797-803.
- Terry R, Palmer S, Rimes K, Clark C, Simmonds J, Horwood J. Living with Joint Hypermobility Syndrome. Patient experiences of diagnosis, referral and self-care. Family Practice 2015; 32(3): 354-8.
- Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LM, de Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology 2007; 60: 34-42.
- Tinkle BT. Issues and management of joint hypermobility. A guide for the Ehlers-Danlos Syndrome Hypermobility Type and the Hypermobility Syndrome. USA: Left Paw Press; 2008.
- Tinkle BT, Bird HA, Grahame R, Lavallee M, Levy HP, Sillence D. The lack of clinical distinction between hypermobility type of Ehlers-Danlos syndrome and the joint hypermobility syndrome. Am J Med Denet A 2009; 149A(11): 2368-70.
- Torrance N, Smith BH, Lee AJ, Aucott L, Cardy A, Bennett MI. Analysing the SF-36 in population-based research. A comparison of methods of statistical approaches using chronic pain as an example. Journal of Evaluation in Clinical Practice 2009; 15(2): 328-34.

- Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). Journal of Manual and Manipulative Therapy 2012; 20(3): 160-6.

Participant characteristics and baseline outcome scores		All participants (n=363)	'Stable' participants (n=233)
Age, mean ± SD (years)		43.8 ± 13.7	44.5 ± 13.9
Sex, women : men (% women)		346 : 17 (95.3%)	221 : 12 (94.8%)
Relationship status, n (%)	Single	100 (27.6%)	71 (30.5%)
	Marrried/ partner	223 (61.6%)	134 (57.5%)
	Divorced/ separated	30 (8.3%)	21 (9.0%)
	Widowed	7 (1.9%)	5 (2.1%)
	Other	2 (0.6%)	2 (0.9%)
Living arrangements, n (%)	Alone	61 (16.9%)	45 (19.4%)
	With husband/ wife/ partner	222 (61.5%)	132 (56.9%)
	With somebody else	78 (21.6%)	55 (23.7%)
Education	Years at school, mean ± SD (years)	12.8 ± 1.5	12.8 ± 1.5
	College diploma or equivalent, n (%)	Yes 203 (67.2%) No 99 (32.8%)	Yes 135 (67.8%) No 64 (32.2%)
	A university degree or equivalent, n (%)	Yes 175 (56.3%) No 136 (43.7%)	Yes 114 (57.9%) No 83 (42.1%)
	A postgraduate degree (e.g. PhD), n (%)	Yes 67 (25.0%) No 201 (75.0%)	Yes 43 (25.1%) No 128 (74.9%)
Paid job at present, n (%)	Yes	195 (54.5%)	127 (55.2%)
	No	163 (45.5%)	103 (44.8%)
Nature of paid job, n (%)	Part-time	88 (48.1%)	59 (50.9%)
	Full-time	95 (51.9%)	57 (49.1%)
	Self-employed	30 (18.9%)	21 (20.4%)
	Employee	129 (81.1%)	82 (79.6%)
No paid job at present, n (%) *	Retired	45 (26.2%)	32 (28.1%)
	Unemployed and seeking work	15 (8.7%)	11 (9.6%)

	Early retired due to sickness or disability	79 (45.9%)	50 (43.9%)
	Full time student	13 (7.6%)	7 (6.1%)
	Doing voluntary work	22 (12.8%)	15 (13.2%)
	At home doing housework	29 (16.9%)	22 (19.3%)
Ethnicity, n (%)	White	344 (95.3%)	216 (93.5%)
	Mixed	6 (1.7%)	6 (2.6%)
	Asian	2 (0.6%)	2 (0.9%)
	Black	2 (0.6%)	2 (0.9%)
	Chinese	1 (0.3%)	1 (0.4%)
	Other	6 (1.7%)	4 (1.7%)
Baseline BIoH score, mean ± SD		228.8 ± 53.8	223.6 ± 54.0
Baseline BIoH score, range		42-326	42-325
Baseline number of body pain areas, mean ± SD		8.0 ± 2.1	7.8 ± 2.1
Baseline SF-36 PCS		32.8 ± 8.9	33.6 ± 9.2
Baseline SF-36 MCS		41.7 ± 11.6	41.9 ± 11.6

Table 1. Characteristics and baseline outcome scores for all participants (n=363) and the subgroup of 'stable' participants (GRC score -1 to +1, n=233). % figures are expressed as a proportion of valid responses to each question. *Responders could choose more than one option so total % may be more than 100% (% is expressed as a proportion of those responding to this question, n=172 and n=114 respectively). MCS = mental component score; PCS = physical component score; SD = standard deviation; SF-36 = Short Form-36 questionnaire.

Criteria for diagnosis		'Yes' n (%)
A. Hakim and Grahame (2003) criteria	**≥2/5 of the criteria below**	**338/363 (93.1%)**
	1. Can you now (or could you ever) place your hands flat on the floor without bending your knees?	312/361 (86.4%)
	2. Can you now (or could you ever) bend your thumb to touch your forearm?	280/363 (77.1%)
	3. As a child did you amuse your friends by contorting your body into strange shapes OR could you do the splits?	262/360 (72.8%)
	4. As a child or teenager did your shoulder or kneecap dislocate on more than one occasion?	137/361 (38.0%)
	5. Do you consider yourself double-jointed?	249/359 (69.4%)
B. Have you received a formal diagnosis (from a healthcare professional) of:	**≥1 of the criteria below**	**351/363 (96.7%)**
	1. Joint Hypermobility Syndrome	286/310 (92.3%)
	2. Ehlers Danlos Syndrome – Hypermobility Type (formerly EDS-III)	189/302 (62.6%)
Met criteria A. and/or B.		**363/363 (100%)**

Table 2. Criteria used to confirm a diagnosis of JHS for all participants (n=363). % figures are expressed as a proportion of valid responses to each question.

Outcome measure	$\bar{d}$ (95% CI)	SD_{diff}	SEM	SDC	95% LoA
BIoH (max 360 points), n=233	-3.26 (-5.98, -0.54)	21.20	14.99	41.55	-44.81, 38.29
SF-36 PCS (max 100 points), n=227	0.00 (-0.57, 0.57)	4.37	3.09	8.56	-8.56, 8.57
SF-36 MCS (max 100 points), n=227	0.55 (-0.48, 1.58)	7.88	5.57	15.44	-14.89, 16.00

Table 3. Calculations related to limits of agreement analysis. BIoH = Bristol Impact of Hypermobility questionnaire; CI = confidence interval; $\bar{d}$ = mean difference scores; LoA = Limits of agreement; MCS = mental component score; PCS = physical component score; SDC = Smallest Detectable Change (calculated as 1.96 x SD_{diff}); SD_{diff} = standard deviation of difference scores (the difference in score between baseline and week 2); SEM = standard error of measurement (calculated as $SD_{diff} \div \sqrt{2}$); SF-36 = Short Form-36 questionnaire.

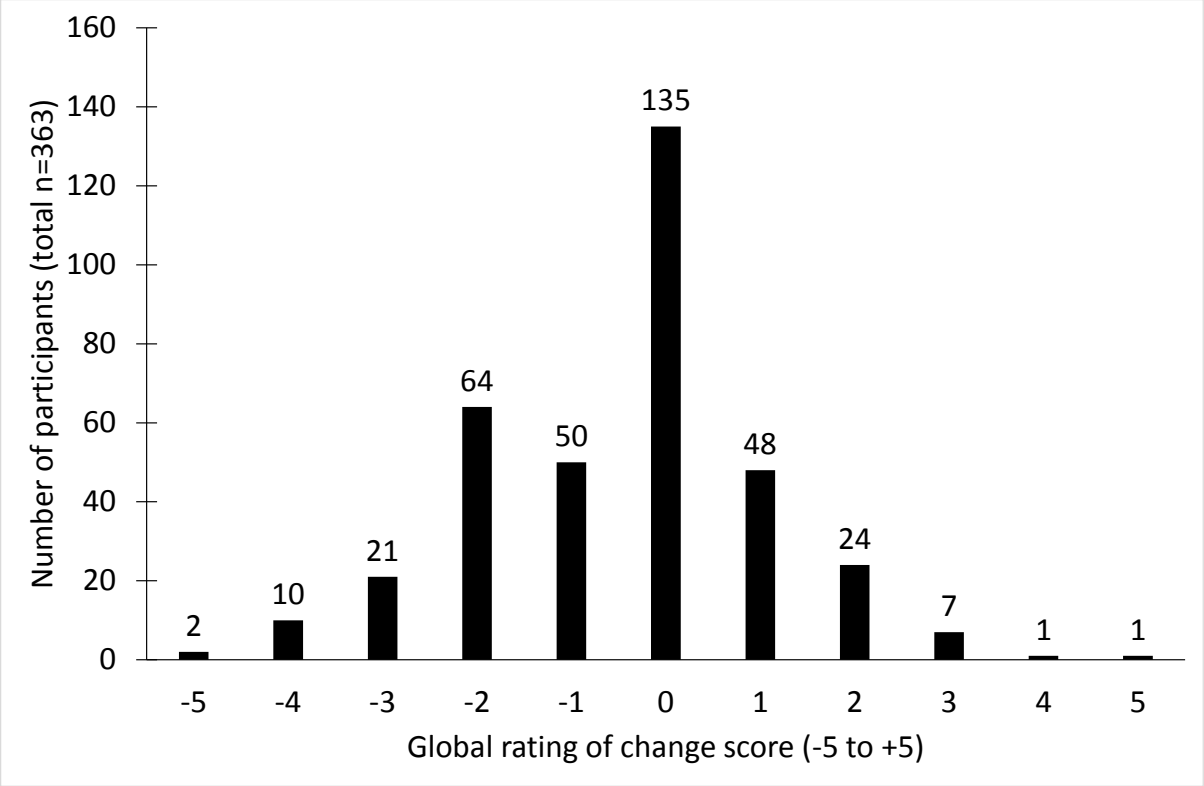

Figure 1. Distribution of global rating of change scores for all participants (n=363).

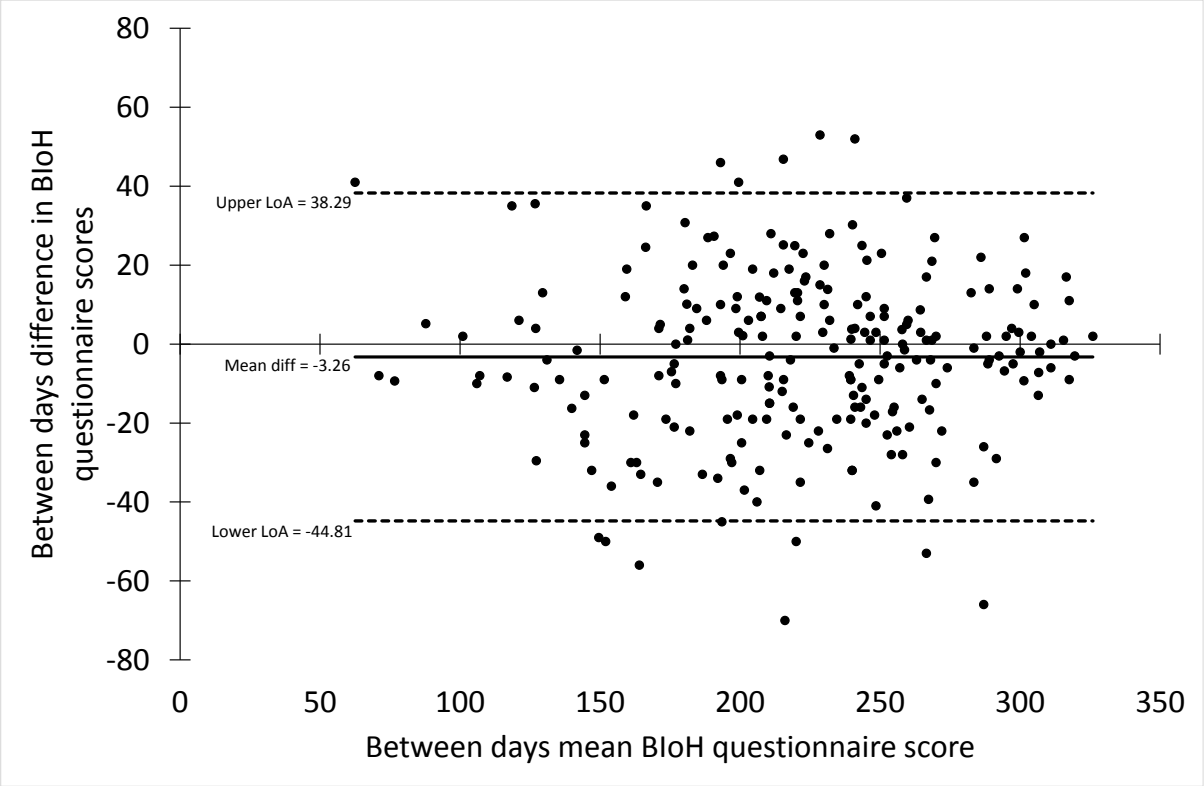

Figure 2. BIoH questionnaire limits of agreement plot for 'stable' participants (GRC score -1 to +1, n=233). Mean between days BIoH scores are plotted against the between days difference in BIoH scores for each individual participant. The superimposed lines depict the mean difference score (solid line) and 95% limits of agreement (dashed lines). LoA = Limit of Agreement. Mean diff = mean difference.

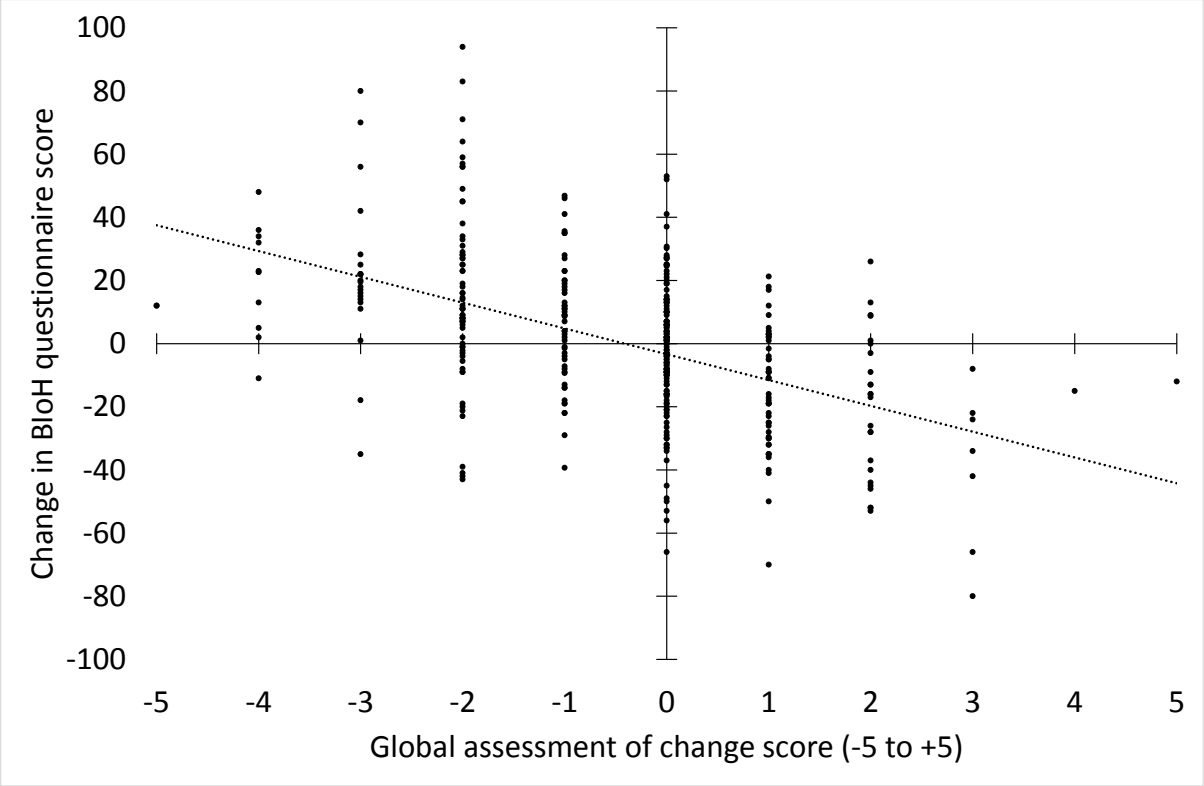

Figure 3. Scatterplot of global assessment of change score against change in BIoH questionnaire score for all participants (n=363).